View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Publications at Bielefeld University

1. So What's Your Affiliation With Gesture?

Carolin Kirchhof [1]

[1] Faculty of Linguistics and Literary Studies, Bielefeld University, Bielefeld, Germany
ckirchhof@uni-bielefeld.de

Abstract

Following De Ruiter (2000), I propose that there is no such thing as a lexical affiliate for every gesture. I suggest interpreting gestures by their *conceptual affiliates*. The existence of conceptual rather than lexical gesture affiliates is supported by empirical data from a perception study with German native speakers. They linked gestures in video clips without sound to their accompanying speech that was in a separate audio clip. The manifold lexical connections people made could be united when regarding them as parts of *conceptual* perceptions. Also, the conceptual affiliates are closely connected with the theme-rheme structure of utterances. This connection further supports the intricate relationship of speech and gesture. The phenomenon of conceptual affiliates implies a less restricted and more natural perception process of co-expressive gesture and speech than the long-held idea of lexical gesture equals.

Index Terms: gesture, meaning, perception, lexical affiliates

2. Introduction: Connectedness of speech and gesture in natural discourse

The fact that gestures and their affiliated speech are produced roughly simultaneously has long been established ([1]-[6]). In contrast to other areas of psycholinguistics, the focus in gesture research has mainly been on production rather than perception (e.g., [1]-[2],[7]-[13]). Although several studies have looked at the audiovisual perception of gestures (e.g., [14]-[16]), the issue of the perceived simultaneity of gestures with speech has rarely been addressed explicitly. While strong asynchrony between the modalities (mismatching) causes the speaker to repair their utterance (e.g.,[2],[17]), the listener is expected to disregard or align the smaller asynchronies internally. A starting point for learning more about the processes of this alignment might be *lexical affiliates*. These parts of utterances that directly correspond in meaning to gestures have also largely been discussed in relation to language production but rarely from the viewpoint of comprehension.

Several researchers have set out to find 'the' lexical item affiliated to a co-occurring gesture (e.g.,[12],[18]). Finding this item seems to be straightforward when people say “Look over there” and simultaneously point at the intended location. It does get more complicated, though, when trying to figure out what Anne meant by “spraying water” down with her right hand while saying “The yard looked so beautiful”. What was she trying to express?!

To investigate how listeners make sense of co-expressive speech and gesture, we did a perception study with German native speakers in which we showed the participants iconic, spontaneous gestures in (soundless) video clips. They then connected the gesture meanings to the co-occurring speech that was played separately. There were no restrictions on the range or order of the speech (words, parts of words, phrases etc.) they could connect with the gesture. The lexical connections the participants made are manifold and at first glance often unrelated. Viewing them as parts of conceptual perceptions or general ideas, however, showed a far more unified picture. Language, or communication, comes from images in our minds. It seems only logical that we also perceive such an image when we pay attention to someone speaking. This paper describes a holistic analytical approach to speech-gesture co-expressiveness that is not necessarily restricted to one lexical affiliate. Neither does the approach separate those gestures synchronized with stressed words in the utterance from the others, which has also been a long tradition in gesture research. This is possible because of the separate presentation of audio and video in the study.

3. Issues of speech-gesture affiliation

The close synchronization of speech and gesture is a phenomenon fundamentally discussed already by McNeill [1]-[2] and Schegloff [19], and later also, among others, by De Ruiter and Wilkins [8] Morrell-Samuels and Krauss [20], and Kendon [32],[7]. Apart from the temporal overlap of speech and gesture, shared meaning has also been a central issue in speech-gesture research. Roughly speaking, gestures co-occurring with speech have been proven to semantically support a so-called *lexical affiliate*. An example for this is pointing when giving directions. The class of iconic gestures, however, is not regarded as being fully synonymous with a word or phrase. These movements imitating, shaping, or generally visualizing words are claimed to add to the meaning of an utterance. How this affiliation is defined varies among researchers. In this paper we will develop a working definition of *gesture-speech affiliation*. We will clarify which part(s) of an utterance a gesture targets on the basis of our empirical data. First we will give a chronological overview of literature, focusing specifically on the experiments conducted by Krauss, Morrell-Samuels, and Colasante [12].

A first definition of the relationship between speech and the gestures accompanying it was formulated in the early 1940s by Efron [20]. On the one hand, the two modalities are strongly connected because gesture can emphasize the content of speech. But gestures are more related to the *how* than to the *what* of an idea and they complement what people say from

that side. Efron's distinction of *baton* (rhythmic hand movements) vs. *ideographic* (illustrating ideas) gestures also fits this finding: Sometimes people emphasize what they say with a pounding fist and sometimes they illustrate it with an iconic gesture. Another, more free relation between speech and gesture was also suggested by Efron [20]. The hand or body movements can convey meaning that is independent from speech and so may be synchronous with it or not. Gestures referring to real, non-metaphorical things (*deictics*) as well as *physiographic* ones are also in this framework. These can be related to worldly things by manual description and are either *iconographic* (depicting form or shape) or *kinetographic* (imitating, cf. embodiment) [20]. A last separate category are standardized, culture-specific gestures (*emblematic* or *symbolic*). Ekman and Friesen [22], based on [20], state that such gestures can be used without speech, such as the Western emblem 'thumbs up' (see also [21]). All of the classifications above are among the groundworks of speech-gesture research. The semantic correlation between speech and gesture goes from independent over co-expressive to redundant [20]. This categorization is in its essence still well established.

The "Kendon continuum" that McNeill [2, p. 37] describes, based on [32], is a more specific explanation of the different levels of gesture-speech dependence: "When used in association with speech [Kendon] noted that gesture serves to represent aspects of meaning in a picture-like or pantomimic manner" [6, p. 104]. He classifies gestures along a continuum according to its relation to speech. One pole is the obligatoriness of speech (co-occurring/co-expressive gestures) and the other is its needlessness because of the hands' "codification" (e.g. *emblems* or sign language) ([4];cf. [2, p. 37]). In general, when people speak they may use gesture, but it is not obligatory. Kendon [32] stated that "[gesticulation] seems to replace what might have been a complex descriptive phrase" in certain cases. This would be positioned around the middle of the continuum.

The term 'gesticulation' for spontaneous speech-accompanying hand movements has since [6] been recognized in the research field. Its definition implies that meaning is expressed through both modalities at a time. Again, finding their ideational connection is still an issue. It turned out that gesture strokes usually precede or end at the peak syllable of an utterance, at the sentence stress [5]. Since then, research has often focused and still does on this rather fixed moment to look for a semantic connection (e.g.,[20],[22]).

Building on this temporal synchrony of peaks, the area where gestures support speech in meaning was expanded to that time span "synchronized with linguistic units" [1, p. 351]. While the focus was loosened from being on emphasizing synchrony only the pretty restrictive idea of 'lexical affiliation' [18] had probably been reaffirmed. In [7], emphasis was put rather on semantic coherence, noting that temporal coincidence "appears to be variable" [p.126]. Further research showed that a gesture stroke usually does not follow the stressed syllable in speech [2]. This again supported the "phonological synchrony rule", as it has been called by De Ruiter [23, p. 29). A study by Nobe [24] demonstrated that already the gesture onset precedes the sentence stress. This again gave more weight to the "rule". These discoveries also gave more support to the hypothesis that speech and gesture are co-expressive and should originate from the same idea unit.

3.1. The rise of the lexical affiliate

At the same time that people investigated speech-gesture synchrony and co-expressivity, a more concrete semiotic relationship between the two modalities was introduced by Schegloff [18] (see also 11]-[12]). Apparently, "various aspects of the talk appear to be 'sources' for gestures affiliated with them" [18, p. 273]. This implies that certain parts of an utterance stand in a more direct relationship with a gesture than the rest of it. McNeill [3, pp. 37f.] gives a comprehensive summary of Schegloff's elaborations on lexical affiliation. He describes a *lexical affiliate* as "the word or words deemed to correspond most closely to a gesture in meaning". We have to note here that gesture and speech are not synonymous in meaning but 'merely' co-expressive. With the help of this correspondence, the rheme (qualifying, newsworthy part) of an utterance could be identified even outside syntactical borders. The semantic gestural counterpart actually "tended to precede the words" [18]. This supported that gestures signaled new content in speech. The findings about the affiliation between gestures and words do not specify if the meaningful part of the gesture is its stroke. But since there are parallels to the prosodic peak this can be assumed.

One could say that a gesture and its lexical affiliate stand in a 1-n relation: a gesture may correspond to one or more lexical items inside an utterance. The context of the utterance does not influence this relationship because the kinship in meaning is (fairly) obvious. The lexical affiliate should trigger the gestural counterpart because of the idea they share. The hands are faster because they do not have to respect syntax as much. This happens again and again whenever a gesture fits a linguistic equivalent. And this is where the interpreting side stops matching the production side: We look for the closest match for a gesture in the speech it synchronizes with. We look for synonymy in words, within sentence boundaries. Here, the two concepts of speech-gesture "semiosis", lexical affiliation and co-expressiveness, have to be set apart clearly. On this, McNeill [3] writes that

> [a] lexical affiliate does not automatically correspond to the co-expressive speech segment. A gesture, including the stroke, may anticipate its lexical affiliate but, at the same time, be synchronized with its co-expressive speech segment. [3, p. 37]

Following [3], lexical affiliates can be regarded as a subset of co-expressive speech (co-ex. sp. $\subset$ lex. aff.). Another important point is that the context is involved in the connection between gesture and co-expressive speech. A combination of speech signals can be part of the shared meaning with gesture. Those signals are quite possibly distributed across the utterance and stand in an n-1 relationship with the gesture. This contrasts with the 1-1 relationship gestures have with their lexical affiliates (v.s.).

The characteristic that gesture-speech co-expression sets the rheme apart from the context is another important distinction from lexical affiliation. This is discussed more broadly in the context of McNeill's growth point theory (e.g., [3, pp. 105ff.]). Finally, the stroke-peak synchrony is not as relevant for co-expression: "[T]he time limit on growth point asynchrony is probably around 1~2 secs., this being the range of immediate attentional focus" [25, Ch. 2.4.1). Gesture and speech can still share meaning, even if they are not (fully) synchronous. From the viewpoint of perception this further supports co-expressivity above direct lexical affiliation. A wider scope of bi-modal expression also helps to find the shared meaning of gesture and speech.

Example 1 is part of a typical Canary Row narration. Someone describes Sylvester the cat dressed up in a bellhop uniform. The extract will illustrate the different associations of speech and gesture we just discussed (bold print indicates stress, square brackets the stroke phase):

so n[e **ro**te mit goldenen **knöp**fen]
such a red one with golden buttons

Example 1: Co-expression vs. lexical affiliate.

The speaker traces the position of the buttons on a double button row in a zig-zag motion. The palms of his clawed hands face the chest. The gesture's closest lexical affiliate in Schegloff's sense would be "knöpfen" because he traces the button positions. In example 1 the gesture indeed begins before and ends with this lexical affiliate. But the indexical "so ne" announces a more detailed description of the uniform. It is the rheme's trigger, so to speak, and the gesture's stroke phase begins with "ne". In this utterance, everything from "rote" to "knöpfen" is the rheme. The gesture that overlaps in time with the speech phrase is fully co-expressive to the conveyed image: the bellhop uniform. Without having the context that Sylvester dresses up in a bellhop uniform, though, the co-expressivity hypothesis would not work. but with "knöpfen", that of lexical affiliation would. Since both speaker and listener will naturally have this context, this is not a problem. They are both in the same communication setting and take in both modalities at the same time. Both can perceive the full image.

Lexical affiliates and gestures' semiotic efficacy was also looked at by Krauss at al.[12]. Shortly after [12], Morrell-Samuels and Krauss said that "the onset of gestures usually precedes the *word* they are affiliated to" ([23, p.342, emphasis added]. This goes against the wider definition of co-expression and narrows down lexical affiliation once again. 2.2 is a summary of the experiments done in [12]. They will become relevant later in this paper for the choice of methodology for the perception study we did (Section 3).

3.2. Krauss, Morrell-Samuels, and Colasante (1991)

Krauss et al. [12] hypothesized that the semantic affiliation of gesture and speech "is a post-hoc construction deriving primarily from the listener-viewer's comprehension of the speech and bears no systematic relation to the movements observed" [12, p. 744]. In other words, they claim that lexical affiliates are somewhat forced interpretations. Of the five experiments examining "the information that conversational hand gestures convey to naive observers" [12, p. 744], two involved recognition memory and will therefore not be discussed here.

Krauss et al. narrowed down the gesture's scope of temporal and semantic synchrony from 'linguistic units' (v.s.) to adjacent words or compounds before conducting their examinations. A group of ten people decided on the lexical affiliates in the stimuli of videotaped photo descriptions before the experiments. This procedure restricted the variation in the perception of the later listeners to a controlled minimum. These subjectively rated affiliate pairs were then mixed with random ones. In the first two perception tests, participants in groups of four chose the lexical item(s) they felt closest to the potential meaning of the accompanying gesture. There was no audio in the stimuli but the lips and facial expressions were visible. This might have made people look for synchrony in the clips. Krauss et al. report that "[f]or 93% of the gestures, a majority of subjects selected the 'correct' lexical affiliate" [12, p. 745, emphasis added]. With this high percentage they want to prove that the agreement between ten people is as reliable as one person's subjective perception. After the two experiments, the researchers grouped the gestural and verbal affiliates into rather squishy semantic categories ('description', 'object', 'action', 'location'). A new analysis of the results now showed a 73% accuracy for actions. This finally lead Krauss et al. to the conclusion that gestures are co-expressive and not fully tantamount or redundant to speech-expressiveness [12, p. 747].

A third experiment they tested whether the "perceived gestural meanings derive mainly from the meaning of their [preselected] lexical affiliates" [12, p. 749]. The semantic categories of the gestures were used as a framework again. This time people should identify them from clips that either did or did not have speech or gesture. In a fourth conditions the decisions are solely based on transcripts. This experiment should also test the compensatory functions of speech and gesture for each other (refuted, e.g., by [25, Fig. 4.4: Givón chart with gesture additions]. This is why in most conditions gesture and/or speech were missing.

Krauss et al. conclude from unclear results that it is rather impossible to measure the contribution of gestures to the meaning of an utterance in percentages. But, they found "the association between the semantic category assigned to the gesture and the semantic category of the lexical affiliate is greater when the coder can hear the sound " [12, p. 750]. The researchers interpreted this finding to mean that speech will give gesture an other interpretation than gesture alone. Others (e.g. [4]) have long called this phenomenon emblematicity, which is the level of the gesture's codification and its dependence on speech (see also section 2 in this paper). Also, the question arises whether co-occurring speech and gesture can actually differ in their semantic category. This problem does disappear when we expand the semiotic focus of a gesture to more than a lexical affiliate.

In the general discussion of their study Krauss et al. concluded that gestures helped resolve ambiguity when no greater context was given. So they did to some extent ascribe communicativeness to gestures[12, p. 751]. The authors also recognize that the content of the two modalities is semiotically related. This relationship was, however, unreliable and imprecise. Finally, Krauss et al. hypothesized on the function of gestures in general. They included communicative intention, the mutual compensatory function with speech, and the helpfulness of gesturing with lexical retrieval [12, p. 752]. On the concept of lexical affiliates, Krauss et al. argued that gestures either preceded or fully synchronized with their affiliate. With this finding they confirmed Schegloff [18]. They specified that the synchrony of gestures with words that were more familiar to the speaker started later than when the words were less familiar to them [12, p. 75]. Finally, the higher frequency of gestures in face-to-face interaction than in picture descriptions calls for further studies with more natural discourse.

To conclude the summary of [12], we will briefly outline the core problems with the lexical affiliation of gestures in the study: Presenting subjects with pre-defined affiliates does not contribute to general assumptions on gesture perception. The unclear semantic categories (see [26] for a general discussion of the subjectivity of semantic fields) and their collision with codification take away the spontaneity aspect of gestures. Since all stimuli involved description of photographs, using this as a category would be sufficient. Also, the visibility of lips and facial expressions in the video stimuli might have influenced the judgments of the participants. Finally, the restriction to gestures and their democratically selected lexical

affiliates excluded the possibility of further co-expressive speech that might be needed to comprehend the stimuli.

We will disprove the general presupposition of definite lexical affiliates (cf. Table 1) as well as further flaws of this rather strict association of gesture and speech in the following experiment. We will demonstrate that conceptual affiliation is a more realistic and cross-subject reliable hypothesis.

4. The lexical affiliate on probation

Ten subjects decided on the lexical affiliates in [12]. They were then the foundation for a set of perception experiments. People had transcripts at hand and the option of discussing the possible affiliations. Also, the decision process for 'the' lexical affiliate for each gesture was driven by an intention to agree. This lead to a somewhat standardized subjective perception in the group (within-group variation was not commented on in [12]). But when we want to test a hypothesis of lexical affiliation, no agreement should be forced or intended from the beginning. Another point to be kept in mind is that temporal synchrony and lexical affiliation often go hand in hand in the interpretation of speech-gesture semiosis. Seeing lip movements and facial expressions in a stimulus, even if sound and video are separated, will lead people to look for a connection. The Krauss et al. experiment did not avoid this issue either.

We designed a study that let subjects observe speech and gestures without obvious synchrony or too strict instructions. This guarantees the full variety or unity in the perception of lexical affiliates. In a previous trial study subjects should identify lexical affiliates from three possible co-occurring sentences after they had watched gesture clips without sound. This turned out to be unsuccessful because (a) the situation felt to unnatural to the subjects and (b) the disctractor sentences were perceived as suitable as the original ones to fit the gesture. These results lead us to further extend the context in the stimuli further. In the present study we showed speech and co-occurring gesture in direct succession. People noted down the lexical items that in their opinion were most connected to the meaning of the gesture.

4.1. Stimuli

We already had a corpus of Canary Row narrations by German native speakers (recorded at Bielefeld University in fall 2010) that contained a fairly standardized set of natural but content-controlled speech. We selected a set of twelve fairly large iconic (imagistic) gestures from a set that had been previously annotated for phases in ELAN [27]. At this stage, the content of the co-occurring speech did not matter because the focus was on the size and vividness of the gestures. The stimuli were produced half by women and half by men and were performed in front of the torsos and heads (central and upper central gesture space, cf. [3, p. 274]). People were videotaped frontally so that all upper limbs are visible at all times. The salience of the movements was captured with high speed cameras (205 fps). We transformed the stroke phases (with some milliseconds around it for smoothness) into silent standalone DivX video clips (MPEG-4 Version 5) in the dimension 640x480 pixel. For this we used the GNU public license software VirtualDub [28]. The video clips have an average duration of 1.83 seconds.

The whole sentences or clauses with sufficient contextual information that were the original co-occurring speech were made into uncompressed wave files (16-bit PCM, 44100 Hz) with the public license software Audacity [29]. The average audio segment is 2.75 seconds long and has one verb in about 8.45 words. This depends on the gesture duration as well as on the information necessary to comprehend the speech. All clip pairs have a similar time relation and all were naturally co-occurring.

One of the clip pairs is example 2. It is the same gesture as in example 1 but this time with sufficient contextual speech. The original gesture phrase includes the preparation and retraction of the stroke (2.688 sec). The cut video clip contains the stroke phase of the gesture and tolerance measures (1.99 sec). The speaker traces the position of buttons on a double button row in a zig-zag motion on his chest with claw hands while his palms face the chest. The accompanying speech in the audio clip includes a breath pause "#" and the unfilled pauses "/" (6.64 sec.). The bold font in the transcript indicates sentence stress, the square brackets the position of the stroke phase. Because of the different lengths of the audio and video clip, people could not directly perceive this arrangement, though.

```
sylvester öffnet die tür #
sylvester opens the door #

in seiner  pagenuniform   /
in   his  bellhop uniform /

so  n[e   rote  mit  goldenen knöpfen]
such a  red one with  golden buttons

und  so   /
and stuff /
```

Example 2: Perception stimulus 1.

No other gesturing happened during this utterance. The speaker simply had his hands folded on the table before and after. People knew that the corresponding clips naturally belonged together and that because of that some kind of connection did exist.

4.2. Experiment design

18 native speakers of German (14 female, 4 male, $\bar{x}$ 26 yrs., 13 right handed, 5 left), either studying or working at Bielefeld University, voluntarily took part our the study. Neither of them had concentration problems and those who needed any wore glasses or contacts. We promised them neither credit points nor financial reward. A researcher supervised the participants at home or in university rooms during the study. People could always ask for specifications and no pressure, such as time or performance requirements, was put on them.

The subjects sat in front of a notebook (1280x800 pixel res.) and wore closed headphones (Sennheiser HD 201). They had mouse control over a folder containing the twelve audio-video file pairs. The 18 participants had two sequential clip orders among them so we could exclude the influence of sequentiality in the study. People could regulate the volume but screen contrast and brightness was constant. This way, we could control sufficient detail and visibility of the gestures. The faces in the stimuli were covered and so the subjects were not able to try reading lips or gaze. People were asked to watch and listen to the clips with corresponding file names (e.g. "01.avi" and "01.wav") as often as they liked. They controlled the frequency themselves with the PC mouse. We told the participants about the source of the clips, namely Canary Row re-tellings, and also, if necessary, explained the general course of events in these cartoons. After watching a stimulus pair, they should note down the word, words,

phrases, or parts of words in either position of the utterance which they thought was or were connected with the gesture in meaning. The pre-numbered form also had these instruction in its header. We chose such a wide range for possible affiliates to let the subjects pick whatever made most sense to them. They also underlined those parts of speech they felt were related in meaning to the gesture to verify their perceptions. Of course they could go through the clips another time before doing that. The average session lasted about 15 minutes.

4.3. Results & Discussion

Up to 14 different affiliate tokens occurred per clip-pair. In this context, by token we mean a word combination or word that does not occur in exactly the same way with another subject. Example 2, for instance, reached this maximum number. On average, 7.75 different affiliate tokens (median=9) were noted down by the 18 participants for one stimulus pair. This variation does include minimal differences such as word parts. For example, “trapezförmig” (trapezoid) and “trapez” (trapeze) or “schwingen” (swing) and “rüberschwingen” (swing across) were counted as different tokens.

In a second step, we formed sets of affiliate types from the tokens. For that we took word stems, optional pronouns, etc. into account. The two example pairs from the previous paragraph would now be in two affiliate types. on occasion, though, same tokens had to be sorted into two types because they included two different aspects, such as action and shape. The sorting resulted in a reduction of differing “lexical affiliates” by about 3 to 4.7 affiliate types per stimulus (median=4). For example, the stimulus with the speech segment “Er lüftet dankend den Hut” (He thankfully lifts the hat) went from 10 affiliate tokens to four affiliate types via this process. The core lexemes of these types were “dankend”, “Hut”, “lüftet”, and “lüftet den Hut”. In this case, we grouped the tokens into these four because the emphasis in the affiliates were either on thankfulness, the object of action, the action, or all at once. This variation is, however, still far from a unison affiliate decision as presented by [12]. This may be due to the subjects making their associations independently in our study.

From the viewpoint of co-expressive speech and the McNeillian imagery-language dialectic [2]-[3] a more homogenous grouping of the subjects' speech-gesture affiliates is still possible: We now considered the *conceptual* overlaps instead of lexical or grammatical commonalities. For instance, the 'hat lifting' we mentioned in the previous paragraph was lexically connected to either the hat or the lifting by a lot of people. The idea that unites them all is the action of lifting the hat – the concept that is both expressed in the speech and in the imitative gesture. We sorted all tokens/types of each stimulus by concept and got an average of 2.75 *conceptual affiliates* (median=2). Table 1 shows a distinct reduction from lexical to conceptual affiliate using example 3.

```
/ und schmeißt eins von diesen /
/ and  throws  one   of  these /

trapezförmigen ähm gewichten
 trapezoid     uhm  weights

a[uf die andere seite] #/
onto the other   side  #/
```

Example 3: Transcript of stimulus 1 (~5 sec).

In example 3, the gesture stroke (~2 sec.) synchronizes with “auf die andere seite”. Both hands in chest height, the palms facing each other chest-wide apart throughout, the fingers fanned a bit – the hands tilt forward and freeze half way to the table. The configuration stays the same through the unfilled pauses and then the hands are folded to rest on the table.

Table 1 shows the subjects' individual perceptions of the gesture's relation to the utterance. The parts they felt to be most related to the meaning of the gesture are listed in alphabetical order in the first column 'What people assigned'. '/' in the table means that someone gave no answer. The second column gives a rough English translation of the first column. The 'different affiliate tokens' in column 3 represents people's answers in a clearer form. Each minimally different entry of column 1 is given its own letter in alphabetical order. The coloring of the cells will help seeing the developments in the table. One person, for example, chose “Gewicht” as related semantically to the gesture in the clip they saw. As the fourth new lexical item it is given (d) in column three. The same label is assigned to all inflections of “Gewicht”, such as its plural “Gewichten”. A different lexeme or a combination of words again is labeled differently (“schmeißt, Gewichte”=(e)). Maybe object and action were linked differently to the gesture. Answers (e) and (h) demonstrate this perceptual difference perfectly as the subjects found both aspects noteworthy. This also results in (e) being a combination (d,j) in the 'affiliate

What people assigned	EN equivalent	different affiliate tokens	affiliate types	conceptual overlaps
/	/	a	a	a
auf die andere Seite	onto the other side	b	b	b
auf die andere Seite	onto the other side	b	b	b
diesen	those	c	c	c
Gewicht	weight	d	d	g
Gewicht(n)	weight(s)	d	d	g
Gewichten	weights	d	d	g
Gewichten	weights	d	d	g
Gewichten	weights	d	d	g
schmeißt, Gewichte	throws, weights	e	d,j	g
Trapez	trapeze	f	g	g
trapezförmig	trapezoid	g	g	g
trapezförmig	trapezoid	g	g	g
trapezförmig	trapezoid	g	g	g
trapezförmigen	trapezoid	g	g	g
trapezförmigen	trapezoid	g	g	g
trapezförmigen Gewichten	trapezoid weights	h	d,g	g
und schmeißt	and throws	j	j	g

Table 1: From lexical to conceptual affiliate in example 3.

type' column. Other subjects chose one side of things only. The variable (b) groups those affiliates relating to position, (d) to the weight, (g) to the weight's shape, and (j) to the action of throwing. (e) cannot be connected to sentence stress, but it was possibly perceived as an emphasizing or beat gesture. Among the 5 (7) affiliate types we can find no lexical agreement that explains a common comprehension of the gesture. How would communication work if we always only agreed on its meaning half of the time?

We do find a large conceptual overlap within all stimuli of this study (rightmost column). While people favored one lexical affiliate over another, the image they perceived and then tried to connect to the utterance was the same: a trapezoid weight, the rheme of the utterance, the newsworthy content (cf., e.g., McNeill 2005). When we take the missing answer (a) out of the calculation we get a conceptual agreement of 82.4%. This is far more than either affiliate token or type could supply. A different grouping of the original (b) with (e) and (j) would still result in a vast majority for the weight. On the other hand, if we took the influence of immediate and wider context as discussed by McNeill [2]-[3] into account, the newsworthy information regarding this episode would be as follows: Sylvester is attempting to get to Tweety with the method "catapult" - the fact that the cat is hunting the bird was established in the instructions. The context given to the subjects in this study was merely that of the general Canary Row scenario, and this episode was either first or last in the collection of twelve stimuli. So, it could contrast with the standard cartoon plot. Then the "auf die andere Seite" would be just as newsworthy as the catapult. Or, the immediate background according to the stimulus sequence would be Tweety's owner beating Sylvester up with her umbrella. So, one could argue for both conceptual affiliates (c) or (g) on the basis of co-expression, newsworthiness, and the restrictiveness of lexical affiliates.

In total, the twelve stimuli had an average conceptual affiliate accuracy of 80.3%. Among the twelve, we found a conceptual agreement rate of 95.88% on average (excluding non-answers). The transcripts of the deviating two samples are shown in examples 2 and 5.

```
so    ne rostige regen[rinne
such  a   rusty  rain spout

 die war   neben] dem fenster
that was  next to the window
```

Example 4: Falsification 1.

We discussed above that a conceptual affiliate goes hand in hand with the rheme of an utterance, or its newsworthy part. Example 4 is faulty in two ways: it is lacking a verb in its theme, or main sentence, and it has no obvious rheme ("regenrinne" as an object and/or the rain spout's position). The subject's gesture is a slightly concave wiggling right hand that moves from central position towards the head. We recognize the "rising hollowness" (cf. [2]), but for the participants the context is not sufficient. The design and position of the gesture are irrelevant without the information that Sylvester is crawling through the pipe. 8 out of 18 subjects could not connect the gesture to the utterance at all, 3 chose the pipe's position and 4 the factual pipe. Also, two people connected the gesture to "so ne" ('such a'), interpreting the gesture as interactional rather than co-expressive. The 30% (position) to 40% (object) agreement of conceptual affiliates is distinct in contrast to the average 95.88% conceptual agreement. The fact that the utterance is not a complete sentence and has two clauses (rhemes) explains the difference in concepts people connected to the gesture. This makes a point for the co-expressiveness of gesture in the context of themes and rhemes.

Example 2 demonstrates a further falsification of the conceptual affiliation of speech and gesture. In contrast to example five, this audio clip does not have a potential lack of themes/rhemes. Instead, there is one too many, namely (1) "Er öffnet die Tür in seiner Pagenuniform" (opening door in uniform) and (2) "so ne rote mit goldenen Knöpfen und so". The two clauses are not only separated by an unfilled pause, they also complement each other. The rheme of (1) is the opening of the door (in uniform) and (2) further specifies (1) with a description of the uniform. The gesture zig-zagging across the chest could have triggered two or even three conceptual affiliates. One is the button design (38.8%) and another the uniform in general (33.3%). 4 subjects also included "öffnen" (opening) in some combination or other (22.2%) and so had a third concept in mind. In contrast to these two cases, all stimuli with only one rheme made a unification of the affiliates that were picked by the participants possible that was between 82.4-100% (median=100).

5. The fall of the lexical affiliate

The study we performed shows that an iconic gesture corresponds to a rheme, and one rheme only at a time. With examples 2 and 5 we have demonstrated that an uneven relationship between gesture an rheme makes people perceive the same utterance in different ways. This phenomenon has been explained through *conceptual affiliation*. A speaker has an idea in their head they want to express. Speech and gesture are used together to convey this idea. While speech is bound by syntax and the lexicon the hands may move freely. The crux of the modalities' co-expressivity was long thought to be their synchrony (especially of stroke and peak). We excluded this factor as well as we could in our study. Since the audio clips in the experiment had more than one prosodic peak in general and only one gesture phase, the subjects simply were not able to connect the two in unison. Also, we let the participants decide independently which gesture-affiliates to pick in the speech without regulations for position or number. This did not force the idea of 'the' lexical affiliate on them but the condition did not exclude it either. The exact choice of words one person made was often also picked by another. But this rarely happened more than twice with one stimulus.

We widened the scope of people's lexical affiliates to include inflections and minor additions such as determiners or pronouns. This already helped with forming larger groups of affiliates. We called the two affiliate groups types and tokens, one being a subgroup of the other. But still the same stimulus seemed to trigger different associations in people.

After we took a closer look at the data we discovered that the instructions to pick words had made people decide for different aspects of the stimuli. For instance, the action of throwing or the object being throw (see Table 1) could be affiliated with different parts of the speech stimulus. What most answers for the stimuli had in common, though, were the connection of the gestures with parts of the utterance's rhemes. For example 3, 14 out of 17 participants noted down that the weight – its existence, shape, or being thrown – was the part of the utterance that was related in meaning to the gesture. Either selected word is part of the same concept that the majority of people perceived from the stimulus. There is a *conceptual affiliate* for the iconic gesture that corresponds to the rheme of the utterance (a weight being thrown). We found

this method successful for all stimuli with exactly one rheme and one gesture. This confirms De Ruiter's [8] conclusion "that gestures do not have lexical affiliates but rather 'conceptual affiliates'" [8, p. 291]. Also, the finding suggests that in a model of co-occurring speech and gesture, the temporal tolerance of 1~2 seconds as suggested by McNeill [25, Ch. 2.4.1] would not disturb comprehension. This effect is currently being tested in another study.

The conceptual affiliation occurs across the borders of adjacent lexemes (see example 2, "rote mit goldenen"). So, "the notion of a conceptual affiliate can also explain the occurrence of the occasional gesture that seems to be related to a single word" [8]. Finally, conceptual affiliation of speech and gesture also supports why Anne in the beginning of this paper combined a gesture like spraying water with "The yard looked so beautiful" - she meant the sun sparkling on the fresh layer of snow. We know this because she mentioned the beautiful winter weather before in the conversation.

All gestures discussed in this paper in the context of lexical affiliation can be positioned on the mandatory-speech pole of Kendon's continuum. Indeed, studies have been carried out testing the recognition factor of emblematic gestures. In [33], for instance, subjects determined the meaning of emblematic and random hand postures, but without contextual speech and only from pictures. A study investigating such codified gestures in the context of conceptual affiliation should result in a high percentage of overlaps between actual lexical and conceptual affiliates within one cultural community. This would be in parallel to the gesture continuum because emblems can often be regarded as word-like. Further studies are also necessary, to investigate multimodal conceptual affiliation in a natural communication setting.

6. Acknowledgments

I would like to thank Jan de Ruiter for initiating this discussion with me, and David McNeill for pointing me in the right direction. Finally, I am most grateful to Sue Duncan for her enthusiastic encouragement.

7. References

[1] D. McNeill, "So you think gestures are nonverbal?", *Psychological Review*, vol. 92(3), pp. 350-371, 1985.

[2] B. Klaus and P. Horn, Robot Vision. Cambridge, MA: MIT Press, 1986.

[3] D. McNeill, *Hand and mind: what gestures reveal about thought*. Chicago, IL: University of Chicago Press, 1992.

[4] D. McNeill, *Gesture and thought*. Chicago, IL: University of Chicago Press, 2005.

[5] A. Kendon, "Some relationships between body motion and speech. An analysis of an example," in *Studies in Dyadic Communication,* A. Siegman and B. Pope, Eds. Elmsford, NY: Pergamon, 1972, pp. 177-210.

[6] A. Kendon, "Gesticulation and speech: two aspects of the process of utterance," in *The Relationship of Verbal and Nonverbal Communication*, MR Key, Ed. The Hague: Mouton, 1980, pp. 207-227.

[7] A. Kendon, *Gesture: Visible Action as Utterance*. Cambridge, UK: CUP, 2004.

[8] J.P. De Ruiter and D.P. Wilkins, "The Synchronisation of Gesture and Speech in Dutch and Arrernte (an Australian Aboriginal language): a cross cultural comparison," presented at the Conférence Oralité et Gestualité, Besançon, France, 1998.

[9] J.P. De Ruiter, "The production of gesture and speech," in *Language and Gesture,* D. McNeill, Ed. Cambridge, UK: CUP, 2000, pp. 284-311.

[10] D. McNeill and S. Duncan, "Growth points in thinking for speaking," in *Language and Gesture,* D. McNeill, Ed. Cambridge, UK: CUP, 2000, pp. 141-61.

[11] J. Holler et al. , "Do iconic gestures really contribute to the semantic information communicated in face-to-face interaction?," *Journal of Nonverbal Behavior*, vol.33, pp. 73-88, 2009.

[12] R.M. Krauss et al., "Lexical gestures and lexical access: A process model," in *Language and Gesture,* D. McNeill, Ed. Cambridge, UK: CUP, 2000, pp. 261-283.

[13] R.M. Krauss et al., "Do conversational hand gestures communicate?", *Journal of Personality and Social Psychology*, vol. 61, pp. 743-754, 1991.

[14] P. Feyereisen, "How do gesture and speech production synchronise?", *Current Psychology Letters: Behaviour, Brain and Cognition,* vol. 23(2), n.p., 2007.

[15] M.W. Alibali et al., "Effects of Visibility between Speaker and Listener on Gesture Production: Some Gestures Are Meant to Be Seen," *Journal of Memory and Language*, vol. 44(2), pp. 169-188, 2001.

[16] W. Fujisaki, and S. Nishida, "Temporal frequency characteristics of synchrony-asynchrony discrimination of audio-visual signals," *Experimental Brain Research*, vol. 166(3-4), pp. 455-464, 2005.

[17] Petrini et al., "Expertise with multisensory events eliminates the effect of biological motion rotation on audiovisual synchrony perception," *Journal of Vision,* 10(5), pp. 1-14, 2010.

[18] S. Goldin-Meadow. "The role of gesture in communication and thinking," *Trends in Cognitive Science*, vol. 3, pp. 419-429, 1999.

[19] E. A. Schegloff, "On some gestures' relation to talk," in *Structures of Social Action. Studies in Conversation Analysis,* J. M. Atkinson and J. Heritage, Eds. Cambridge: Cambridge University Press, pp. 266-296, 1984.

[20] D. Efron, *Gesture, race, and culture*. Hague: Mouton, 1972. (Original work published as D. Efron, *Gesture and environment*. New York: King's Crown Press, 1941)

[21] P. Morrell-Samuels and R.M. Krauss, "Word familiarity predicts temporal asynchrony of hand gestures and speech," *Journal of Experimental Psychology: Learning, Memory and Cognition,* vol. 18, pp. 615-623, 1992.

[22] P. Ekman and W. Friesen, "The repertoire of nonverbal behavior: Categories, origins, usage, and coding," *Semiotica,* vol. 1, pp. 49-98, 1969.

[23] E. McClave, "Gestural beats: The rhythm hypothesis," *Journal of Psycholinguistic Research*, vol. 23, pp. 45-66 , 1994.

[24] J.P. De Ruiter, "Gesture and speech production," Doctoral dissertation, Catholic University of Nijmegen. Nijmegen, the Netherlands, 1998.

[25] S. Nobe, "Representational gestures, cognitive rhythms, and acoustic aspects of speech: a network/threshold model of gesture production", Doctoral dissertation, University of Chicago. Chicago, IL, 1996.

[26] D. McNeill, *How language began*, unpublished.

[27] C. Kirchhof, "The Truth about Mid-Life Singles in the USA: A Corpus-Based Analysis of Printed Personal Advertisements," Master's thesis, Bielefeld University, Bielefeld, Germany, 2010.

[28] O. Crasborn and H. Sloetjes, "Enhanced ELAN functionality for sign language corpora," presented at LREC 2008, Sixth International Conference on Language Resources and Evaluation, Marrakesh, Morocco, 2008.

[29] A. Lee. (2010). *VirtualDub* (1.9.11) [video editing software]. Available : http://virtualdub.sourceforge.net/

[30] The Audacity Team. (2011). *Audacity* (1.3.13 Beta) [audio editing software]. Available: http://audacity.sourceforge.net/

[31] A. Kendon, "How gestures can become like words," in *Crosscultural Perspectives in Nonverbal Communication*, F. Poyatos, Ed. Toronto: C. J. Hogrefe, Publishers, 1988, pp. 131-141.

[32] T.C. Gunter and P. Bach, "Communicating hands: ERPs elicited by meaningful symbolic hand postures," *Neuroscience Letters* (372)1-2, pp. 52+, 2004.